

## 英特尔® 固态硬盘助力金山云\* 构建 KingStore\* 超融合架构 解决方案



“金山云 KingStore\* 超融合存储搭载先进的英特尔® 至强® 处理器 E5-2600 系列、采用 NVMe\* 的固态硬盘和万兆网卡等产品，为客户提供了高可靠、高性能的解决方案。在整个产品选型中，充分验证了英特尔产品具备低延迟、高带宽、高 IOPS、稳定性好等特性。相信随着双方的合作更加深入，未来也一定会为客户带来更领先的产品以及解决方案。”

北京金山云网络技术有限公司  
合伙人&副总裁  
宋伟

### 要点综述

金山云\* 是中国最大的云存储服务商，为用户提供领先的公有云与私有云服务。为了解决传统存储架构在性能、扩展性、成本效益等方面的问题，金山云将超融合作为未来重要发展方向，并与英特尔（中国）云计算创新中心进行紧密配合，推出搭载了英特尔® 固态硬盘并采用超融合架构的金山云 KingStore\*，这种新型的软件定义存储方案实现了存储容量和性能线性扩展、快速 IO 吞吐、全局负载均衡等功能，可为金融、能源、互联网等行业客户提供高可用、高敏捷性的云服务，推动业务的云化进程。

### 简介

随着大数据、物联网、移动互联等领域最新技术的融合发展，IT 基础架构正经历着向云计算演进的巨大技术变革，大量用户都采用云计算、虚拟化技术来提升资源利用率、增强 IT 敏捷性，这给存储系统的 IOPS 性能、扩展性、成本效益带来了巨大挑战。在此背景下，基于软件定义存储、软件定义计算、商用硬件和统一管理界面来提供存储和计算资源的超融合架构得到了快速的发展。IDC\* 数据显示，2015 年第四季度，全球超融合产品销售额同比增长 170.5%，并有望在未来实现持续高速增长。在此背景下，金山云采用英特尔固态硬盘+英特尔® 处理器+英特尔® 网卡的组合，基于最新的软件定义存储 KingStore 打造了超融合的云平台。

### 挑战

在为用户提供云服务的过程中，金山云发现，如果依然沿用传统的集中式存储架构，将很难实现预定目标，整个平台将会面临如下几个方面的巨大挑战：

- **如何提升 IOPS 能力以满足云计算需求：**基于云计算技术构建的新型数据中心，对整个存储系统提出了极高的要求：具备快速响应、高 IO 处理、高可靠、低成本、线性扩展等特性，以解决数据快速增长带来的挑战。
- **如何提升扩展能力：**考虑到应用系统未来 3~5 年的业务发展，以及业务突发的需求，为满足应用系统的性能、容量承载需求，就需要提升存储系统的扩展性。这不仅仅是要求存储系统能够充分利用现阶段存储设备的硬件能力，不造成浪费，还要求存储系统在进行扩展的时候，不对现有应用的连续性构成影响，并最好能够实现存储容量和性能可线性扩展。



- **提升存储系统的成本效益:** 对于金山云来说, 成本是其考虑的一个重要因素。如果采用高端磁盘阵列的方式, 将会产生极高的硬件购置费用, 带来巨大的成本压力。而目前国内云服务市场竞争极为激烈, 若无法有效控制存储系统的建设成本, 将会极大的制约金山云 KingStore 的整体市场竞争力。
- **减低系统运维难度:** 专属的存储系统往往意味着较高的系统复杂度与管理难度, 这导致数据中心运维效率低下、投入成本高等问题, 难以适应云计算大数据时代对 IT 基础设施的要求。此外, 在数据读写高负载环境下, 也需要尽量避免存储设备产生故障, 并能够在故障发生后尽快恢复数据, 减少对于业务连续性的影响。

### 解决方案

集中式存储解决方案的固有特性并不能解决存储系统在性能、扩展性等方面的瓶颈。因此, 金山云决定开发多节点网络聚合, 实现模块化无缝横向扩展的软件定义存储方案 KingStore, 并采用超融合架构作为实施部署方案。

金山云 KingStore 通过统一管理平台, 实现了对于硬件设备层、存储架构层、存储服务层、存储驱动层的统一管理。在最底层的硬件设备层, 采用了标准的 X86 企业级服务器作为超融合节点, 每个节点搭载英特尔® 固态硬盘 DC S3710 + 英特尔® 至强® 处理器 E5-2680 V4 + 英特尔® 万兆网卡 X520-SR2 的组合。

在存储架构层, 金山云 KingStore 则采用了分布式文件系统 KDFS, 其采用一致性哈希算法, 增强了容错性和扩展性, 可以让数据尽可能分布到所有的节点中, 这样可以使得所有节点自动负载均衡。此外, 其还满足了高效性要求, 当有新节点加入系统中, 系统会重新做数据分配, 数据迁移仅涉及新增节点, 现有节点上的数据不需要做很大调整。

存储服务层包含快照、克隆、压缩、备份、容灾等功能, 可以为用户提供高性能、高效率的存储服务, 同时能够在发生意外事件

时及时恢复数据, 确保业务的高可用性。而存储驱动层则包含块存储服务、文件存储服务、对象存储服务, 可为虚拟化顶层的 OpenStack\*、KVM 等虚拟化系统提供存储能力支撑。

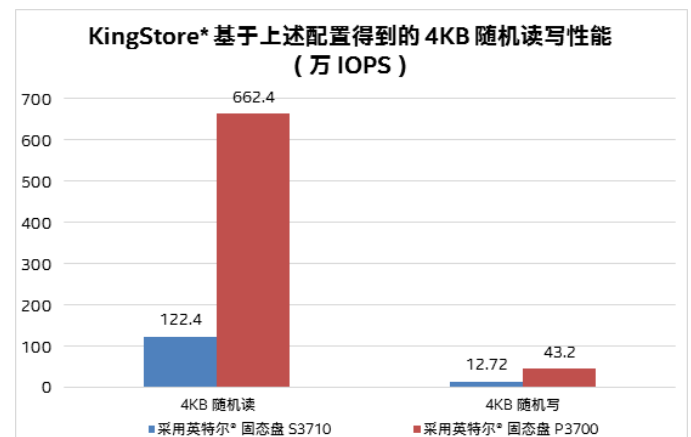
先进的架构、创新的技术确保了金山云 KingStore 能够提供领先的性能、广泛的兼容性、极高的可用性, 能够为行业用户提供高价值的云服务, 并降低云服务的总拥有成本。

### 实施效果

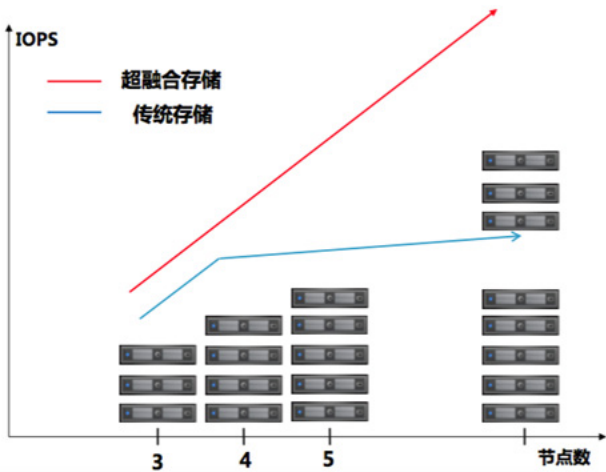
#### 金山云 KingStore 如何实现领先业界的 IO 性能

金山云与英特尔 (中国) 云计算创新中心进行紧密配合, KingStore 采用的英特尔® 固态硬盘 S3710 或 P3700 在硬件本身即具备极高的 IO 能力。在模拟真实环境的测试中 (服务器、硬盘部署于同一个 RACK, 测试配套见下表), 采用基于 SATA 协议的英特尔固态硬盘 S3710 的金山云 KingStore 最终得到平台 4KB 随机读和随机写性能分别为 122.4 万 IOPS 和 18.72 万 IOPS。如果采用基于 NVMe 协议的英特尔 PCIe\* 固态硬盘, 金山云 KingStore 的性能还将有大幅增长。在实际测试中, 采用英特尔固态硬盘 P3700 的金山云 KingStore, 最终测得 4KB 随机读和随机写能力分别为 662.4 万/43.2 万 IOPS, 远超过采用 SATA 固态硬盘的性能, 适用于峰值数据负载量大, 对于业务延迟有着严格要求的客户。此外, 英特尔 SATA 和 NVMe 固态硬盘都具有非常优秀的性能一致性, IOPS 能力随应用环境变化的抖动很小。如此卓越的数据读写性能, 保证了云平台即使在高负载的情况下, 也能向客户提供高速稳定的云服务。

名称	配置描述	数量
服务器	2*E5-2680v4; 128GB Memory; 2*X520-SR2 10GbE NIC	6
硬盘	英特尔® 固态硬盘 DC S3710 400GB 或英特尔® 固态硬盘 DC P3700 400GB	24



值得一提的是，对于传统存储架构来说，IOPS 性能不能随着节点数的增加而线性增长，如果企业想要提高 IOPS 性能，就必须在硬件上耗费大量的成本。金山云 KingStore 采用无中心节点的全对称架构，无单点故障，存储容量和性能可线性扩展，并且支持平滑扩容，容量几乎不受限制，充分保障了存储系统的灵活性和可扩展性。



### 使用英特尔® 固态硬盘确保云服务的高可用性

对于行业用户来说，云服务的可用性至关重要，英特尔固态硬盘帮助金山云 KingStore 更容易接近了这个目标。即使在高负载的情况下，英特尔固态硬盘也能够保证极低的故障率。统计显示，英特尔固态硬盘在 6 百万片出货量中，真实年失效率低于 0.1%，这不仅保证了存储系统的稳定性，也极大的降低了运维成本。

此外，英特尔固态硬盘还对可靠性进行了改进，通过冗余存储空间保护用户数据，并提供断电数据保护功能，支持在系统突然断电时利用电容电量将缓存中的数据立即保存后再停止工作。即使固态硬盘发生故障，或者云数据中心出现供电故障也不会影响数据的完整性。在实际测试中，金山云 KingStore 搭载了两套 VMware\* 系统（VMware 系统环境为：CPU=2Core，Memory=4GB；/dev/vda=40GB，/dev/vdb=100GB），并通过插拔固态硬盘模拟了固态硬盘损坏后的数据丢失与恢复情况。

测试结果表明，在固态硬盘拔出并重新插入的过程中，金山云 KingStore 所在 VMware 中的数据写入会在暂停之后恢复，这个过程中数据不会丢失。另外，在固态硬盘损坏或拔出之后，金山

云 KingStore 存储集群会以 20MB/S 的速度恢复数据，这个过程中存储系统的整体 IOPS 性能基本不受影响。

另外，金山云 KingStore 采用集群管理方式，不存在单点故障，灵活配置多数据副本，不同数据副本存放在不同的节点上，单个物理节点故障不影响业务的使用，并通过纠删码技术来提高可用性（纠删码是一种在分布式文件系统中保证数据可靠性的技术，能提供优化的数据冗余度，以防止数据丢失，具有编码效率高、节省存储空间的特点）。

### 存储系统总拥有成本（TCO）得到有效控制

在云存储系统的建设中，存储系统的硬件采购是一笔庞大的支出，相对于昂贵的传统存储架构，金山云 KingStore 通过采用标准的企业级 X86 服务器，整合了计算和存储功能，极大地降低了存储系统的成本支出，借助超融合节点上的英特尔固态硬盘，金山云 KingStore 无需部署昂贵的磁盘阵列就可以获得同等的 IOPS 能力。数据显示，金山云 KingStore 超融合架构单 GB 存储成本只有传统存储的三分之一，从根本上缓解了企业客户的 IT 成本压力。

另外，标准的企业级 X86 服务器还让 IT 人员的运维压力大大降低，企业可以很轻松地将标准化的运维流程迁移到云平台的运维之中。金山云 KingStore 还提供了全图形化的统一操作平台，支持实时性能监控、集群管理、预警及告警机制，充分简化运维人员的操作，运维成本也随之降低。

### 未来规划

英特尔与金山云的强强联合打造了具有极高竞争力的云平台，双方还将在更多层面深化产品合作。除了根据用户需求，敏捷扩展英特尔 NVMe 固态硬盘等更高性能的硬件设备之外，金山云 KingStore 还将集成英特尔® Storage Performance Development Kit（英特尔® SPDK），其提供了一套完整的工具和库，旨在提供高性能、可扩展、高可靠性的用户态存储应用。英特尔 SPDK 具备全用户态驱动，采用轮询模式完成数据收发、避免中断和上下文切换，精确到 per core 的资源配置力度以达到无锁化操作等优势，可以实现资源独占、提高内存使用效率，进而提升 IO 性能。金山云对英特尔 SPDK 提供的 iSCSI Target 与 TGTD 提供的 iSCSI Target 进行了对比测试，测试配置及数据如下：

## 配置:

存储节点配置	3 个节点, 每个节点 12 块 SSD
计算节点配置	1 个计算节点, 配置 5 个 gateway, 每个 gateway 做为一个 iSCSI target 节点, SPDK 仅使用了一个 Core

## 测试数据:

SPDK	4KB 随机写 ( IOPS )	4KB 随机读 ( IOPS )	1MB 顺序写 ( 带宽 )	1MB 顺序读 ( 带宽 )
1个vdi	29835	46835	194	1120
2个vdi	44815 ( 提升了35% )	67487 ( 提升了90% )	319	1121
3个vdi	47551 ( 提升了60% )	65711 ( 提升了84% )	350	1136
4个vdi	58270 ( 提升了82% )	57048 ( 提升了55% )	363	1157
5个vdi	54096 ( 提升了60% )	52528 ( 提升了45% )	374	1142

从测试结果可以看出, 在集成了英特尔 SPDK 之后, 平台的 4KB 数据读写性能将提升 35%-90% 之间, 效果显著。

值得一提的是, 金山云 KingStore 在未来还将整合英特尔® Intelligent Storage Acceleration Library ( 英特尔® ISA-L ) 算法库, 其包括哈希、循环冗余校验、加解密、压缩、纠删码等函数。主要用于加速存储在数据安全、数据保护、数据去重等方面的应用, 英特尔 ISA-L 使用 multi-buffer 技术对哈希进行了优化, 与 openssl 1.0.2 中所使用的哈希对比, ISA-L 哈希最高可达到 15 倍的性能提升。

众所周知, 在分布式存储中, 纠删码是比多备份更高效的保证数据可靠性的技术。纠删码能提供优化的数据冗余度, 以防止数据丢失, 恰当地使用纠删码可以提高空间的利用效率并获得较好的数据保护效果。纠删码具有编码效率高、节省存储空间的特点。但是纠删码的复杂计算导致性能下降。英特尔 ISA-L 提供了高性能纠删码的方案。对于 Reed-Solomon ( RS ) 纠删码 10+4, 在英特尔® CPU 的单核上, 英特尔 ISA-L 纠删码最高可达到 9.7GB/s 的吞吐量, 比目前开源的 RS 纠删码快 10 倍左右。



本文及所提供之信息旨在方便英特尔客户, 并以“概不保证”方式提供, 英特尔不做任何明确或隐含的担保, 包括对适销性、不侵权性, 以及适用于特定用途的担保。收到或拥有本文件, 不代表英特尔对本文所述、显示或包含的内容授予任何知识产权。英特尔® 产品并非设计用于医疗、救生、维生、临界控制、安全防护系统或核设施应用领域。

英特尔公司 © 2016 年版权所有。保留所有权利。英特尔、英特尔标识和至强是英特尔公司在美国和/或其他国家 ( 地区 ) 的商标。

\*文中涉及的其他名称及商标属于各自所有者资产。